

AD-A103 873

WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER
MULTIPLE INTEGRAL EXPANSIONS FOR NONLINEAR FILTERING. (U)

F/G 12/1

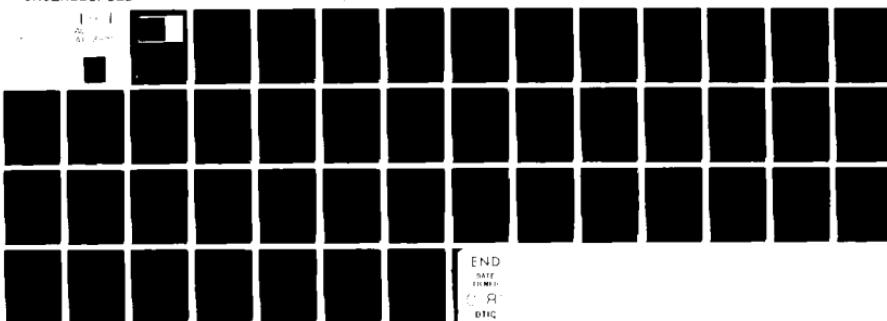
JUN 81 D OCONE

DAAG29-80-C-0041

UNCLASSIFIED

MRC-TSR-2226

NL



END
DATE
TELETYPE
CRA
DTIC

AD A103873

MRC Technical Summary Report # 2226

MULTIPLE INTEGRAL EXPANSIONS
FOR NONLINEAR FILTERING

Daniel Ocone

Mathematics Research Center
University of Wisconsin-Madison
610 Walnut Street
Madison, Wisconsin 53706

June 1981

(Received March 4, 1981)

DTIC
S 1981
A

Approved for public release
Distribution unlimited

FILE COPY

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

Air Force Office of
Scientific Research
Washington, D.C. 20332

National Science Foundation
Washington, D.C. 20550

81 9 08 027

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

Multiple Integral Expansions for Nonlinear Filtering*

Daniel Ocone

Technical Summary Report #2226

June 1981

ABSTRACT

Multiple stochastic integral expansions are applied to the problem of filtering a signal observed in additive noise. It is shown that the optimal mean-square estimate may be represented as a ratio of two multiple integral series. A formula for expanding the product of two multiple integrals is developed and applied to deriving equations for the kernels of best, finite expansion approximations to the optimal filter. These equations are studied in detail in the quadratic case.

AMS (MOS) Subject Classifications: 93E10, 93E11, 60H05

Key Words: Multiple stochastic integral, nonlinear filtering

Work Unit Number 4 - Statistics and Probability

Accession For	NTIS GRA&I DTIC TAB Unannounced Justification	
By		
Distribution/		
Availability Codes		
Print and/or		
Dist / Special		
<i>t</i>		

* This work formed part of the author's Ph.D. thesis in Applied Mathematics at M.I.T. under the supervision of Professor Sanjoy K. Mitter.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041. This material is based upon work supported by the National Science Foundation under Grant No. MCS-7927062, by a National Science Foundation Graduate Fellowship, and by the Air Force Office of Scientific Research under Grant No. AFOSR-77-3281B.

- 1 -

201

SIGNIFICANCE AND EXPLANATION

A common problem in the analysis of stochastic systems is the estimation of a stochastic process given only noise-corrupted or incomplete observations. Examples occur in communications theory when one wants to estimate a signal sent over a noisy channel or in time series problems. If $x(t)$ is a stochastic process denoting the signal, the observations are typically modelled by

$$y(t) = \int_0^t h(x(s))ds + dW(t) ,$$

where $W(t)$ is an independent increments "noise" process, usually Brownian motion. The problem of filtering is to build an estimate, i.e., filter, of $x(t)$ using the observations $y(s)$, $s < t$. Theoretical characterizations of best mean-square estimates are known, but can be translated into effective solutions only in special instances. In this paper, the general filtering problem is treated by attempting to expand filters in series of multiple stochastic integrals of the form

$$\int_0^t \int_0^{s_1} \cdots \int_0^{s_r} a(t, s_1, \dots, s_r) dy(s_r) \cdots dy(s_1) .$$

Two primary issues raised by this idea are considered; representation of the optimal mean-square estimate by multiple integral expansions, and construction of suboptimal estimates using a finite number of multiple integrals. It is shown that expansion of the optimal filter is indeed possible, and a method is presented for finding best, finite expansion estimates. A rudimentary algebra of multiple integral expansions is first developed as a tool to prove these results.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the author of this report.

MULTIPLE INTEGRAL EXPANSIONS FOR NONLINEAR FILTERING*

Daniel Ocone

1.1 Introduction

In the additive noise model of filtering, information about a stochastic process $x(t)$, $t \geq 0$, called the signal, is received through observations of the form

$$y(t) = \int_0^t h(x(s))ds + w(t) \quad t > 0$$

$w(t)$ is a noise term that corrupts the signal, and it is usually assumed to be a Brownian motion. The filtering problem is to estimate from the observations $y(s)$, $0 \leq s \leq t$, a given moment $f(x(t))$ of the signal at time t , and, if estimators minimizing mean-square-error are desired, this means calculating the conditional mean $E\{f(x(t)) | F_t^Y\}$, $F_t^Y := \sigma\{y(s) | 0 \leq s \leq t\}$. $E\{f(x(t)) | F_t^Y\}$ is henceforth referred to as the optimal filter. Two fundamental characterizations of the optimal filter are available: a) a Bayes formula for $E\{f(x(t)) | F_t^Y\}$ as the ratio of two conditional, functional integrals (Kallianpur, Striebel [9], cf. §1.2 of this paper); b), in the case that $x(t)$ is Markov, a representation of the optimal filter as a stochastic integral against the innovations process, $v(t) = y(t) - \int_0^t E\{h(x(s)) | F_s^Y\}ds$, the stochastic integrand being adapted to the observation process (Fujisaki, Kallianpur, and Kunita [2]). However, though theoretically deep, these results lead to explicit and analytically computable solutions only in special instances.

* This work formed part of the author's Ph.D. thesis in Applied Mathematics at M.I.T. under the supervision of Professor Sanjoy K. Mitter.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041. This material is based upon work supported by the National Science Foundation under Grant No. MCS-7927062, by a National Science Foundation Graduate Fellowship, and by the Air Force Office of Scientific Research under Grant No. AFOSR-77-3291B.

This paper studies the application of multiple stochastic integral expansions to the filtering problem. Any filter, optimal or suboptimal is actually an anticipating functional of the observation process, thus suggesting that filters be represented and analyzed within a framework for functional expansions. Multiple stochastic integrals prove useful for this purpose. In fact, their definition originates in Wiener's homogeneous chaos theory, which constructs orthogonal decompositions of spaces of finite-variance functionals of Gaussian processes (cf. Kallianpur [8] and Hida [5]). In the Brownian motion case, each subspace of the decomposition corresponds to the space of multiple stochastic integrals of a given order, and, thus, Wiener's theory shows that any L^2 -functional of the Brownian motion may be expanded in a series of multiple integrals. Multiple integrals have been used already to solve a number of specific estimation problems. Marcus, Mitter, and Ocone [13] apply the homogeneous chaos theory to compute conditional statistics of polynomial functionals of a Gauss-Markov process observed in white noise, and Hida and Kallianpur [6] use multiple integrals to predict non-linear functions of Brownian signals given perfect observations. In cumulant approximations of the conditional density in filtering, Eterno [1] also derives expressions using multiple integrals. Here, we seek to apply multiple integrals of the form

$$\int_0^t \int_0^{s_1} \int_0^{s_{n-1}} a(t, s_1, \dots, s_n) dy(s_n) \cdots dy(s_1),$$

where $a(\dots)$ is deterministic, to the general filtering problem. We focus on two basic issues: the expansion of the optimal filter by expressions involving multiple integrals, and the construction of best suboptimal filters having a finite multiple integral expansion of specified order. It is important to observe that the stochastic integrals we employ are formed from the observation process and not the innovations process. At first, integration against innovations might appear to be an attractive idea because the innovations process is Brownian, integrals of different orders are thus orthogonal, and homogeneous chaos theory can be applied. However, in practice the innovations process is not available since its construction requires the estimate $E\{h(x(t)) | F_t^Y\}$, to compute which is generally a

difficult filtering problem itself. Integrals using $y(\cdot)$ directly are thus more natural, but, due to their more general, usually non-Gaussian character are more difficult to apply. For example, in suboptimal estimation one might like to project random variables on a sum of spaces of multiple integrals. This is easily done for Brownian integrals, using the orthogonality of different order integrals and explicit formulae to calculate the integrands, but not so easily for more general integrals, where the orthogonality structure and kernel formulae are lost. In this paper we describe a method for analyzing $y(\cdot)$ -based integrals, that, in particular, allows resolution of this projection problem.

The paper is organized as follows. §1.2 introduces the precise filtering model we consider and recalls the Kallianpur-Striebel formula for the optimal estimate. A central feature of this formula is the fact that the $y(\cdot)$ process is absolutely continuous with respect to Brownian motion. Transformations of measure so that $y(\cdot)$ becomes Brownian will be an underlying component of our analysis of $y(\cdot)$ -based integrals. §2 is a self-contained treatment of multiple integrals of Brownian and observation processes. We define multiple stochastic integrals, prove technical lemmas for later use, and develop some useful properties of the integrals. Of particular importance is the multiplication formula (theorem 2.1), which shows how to express the product of multiple integrals in a multiple integral expansion, thus providing a rudimentary algebra for handling expansions. We present the applications to filtering in section 3. In §3.1, we show that the optimal filter can be represented as the ratio of two multiple integral expansions, essentially by expanding the Kallianpur-Striebel formula. §3.2 addresses the issue of finding the best (mean square) estimate of the form

$$a_0(t) + \int_0^t a_1(t,\sigma) dy(\sigma) + \dots + \int_0^t \int_0^{s_1} \int_0^{s_{r-1}} a_r(t, s_1, \dots, s_r) dy(s_r) \dots dy(s_1) .$$

By combining the expansions of §3.1 and the multiplication formula, we derive a system of linear integral equations for the kernels $\{a_i\}_{i=0}$. In effect, the method of analysis is to transform measures to a space on which $y(\cdot)$ is a Brownian process and then to apply

the multiplication formula to discover the effect of the Radon-Nikodym derivative so introduced. The remaining sections apply these results, first to rederiving the Kalman filter, second to finding best quadratic filters.

It is a pleasure to thank Professor S. K. Mitter, for suggesting this problem and for inspiring and guiding the research.

1.2 Filtering preliminaries

The precise filtering model to be considered is as follows. Let the underlying probability space be denoted (Ω, \mathcal{F}, P) . For $0 < T < \infty$, let $\{x(t) | t \in [0, T]\}$ be a measurable, real-valued process on (Ω, \mathcal{F}, P) , $h(s, x)$ a Borel function on $[0, T] \times \mathbb{R}$, and $w(t)$ a standard Brownian motion on (Ω, \mathcal{F}, P) , such that

- i) $w(\cdot)$ is independent of $x(\cdot)$
 - ii) $E \int_0^T h^2(s, x(s)) ds < \infty$.
- (1.1)

Set

$$y(t) = \int_0^t h(s, x(s)) ds + w(t) \quad t \in [0, T].$$

Such a process $y(\cdot)$ will be called an observation semi martingale.

Let $f(t; x(s), s \leq t)$ be a non-anticipating functional of $x(\cdot)$ such that $E f(t; x(s), s \leq t) < \infty$, $\forall t \in [0, T]$, and define $P_t^y := \sigma\{y(s) | 0 < s \leq t\}$ and $P_t^{x,y} := \sigma\{x(s), y(s) | 0 < s \leq t\}$.

Theorem 1.1 (Kallianpur, Striebel [9]). Let

$$\frac{dP_0}{dP} = \exp \left\{ - \int_0^T h(s, x(s)) dw(s) - \frac{1}{2} \int_0^T h^2(s, x(s)) ds \right\}.$$

Then (i) P_0 is a probability measure, and P and P_0 are mutually absolutely continuous.

$$(ii) E_0 \left(\frac{dP}{dP_0} \mid \mathcal{F}_t^Y \right) = \exp \left[\int_0^t h(s, x(s)) dy(s) - \frac{1}{2} \int_0^t h^2(s, x(s)) ds \right].$$

(iii) On (Ω, P_0) , $y(\cdot)$ is a Brownian motion independent of $x(\cdot)$.

(iv) $x(\cdot)$ has the same law on (Ω, P_0) as on (Ω, P) .

(v) $E[f(t; x(s), s < t) \mid \mathcal{F}_t^Y]$

$$= \frac{E_0 \{ f(\cdot; x(s), s < t) \frac{dP}{dP_0} \mid \mathcal{F}_t^Y \}}{E_0 \{ \frac{dP}{dP_0} \mid \mathcal{F}_t^Y \}}. \quad (1.2)$$

For a nice treatment of this theorem, see Wong [21]. It is the principal theoretical tool for our work in filtering, for it explicitly characterizes the optimal filter as a functional integral and it establishes that $y(\cdot)$ is mutually absolutely continuous with Brownian motion.

Finally, we remark that we restrict ourselves here to scalar processes only in the interests of notational simplicity. The techniques to be discussed extend easily to the vector case.

2. Multiple Integrals

2.1 Definitions

The concept of a multiple Wiener integral derives ultimately from Wiener's work on 'homogeneous chaos' decompositions of functionals of Brownian motion; however, the modern definition and theory are due to Ito [7]. Here we will define multiple integrals by iteration of stochastic integration. While this differs from Ito's construction, it leads, as Ito [7] notes, to the same result modulo a multiplicative constant. The iterative definition is convenient for our calculations.

Let $(b(t), \mathcal{F}_t^b)$ be a standard Wiener process with its associated family of σ -algebras $\mathcal{F}_t^b = \sigma(b(s) : s < t)$. Recall that, for a jointly measurable, \mathcal{F}_t^b -adapted process $\phi(t, w)$ such that $E \int_0^T \phi^2(s) ds < \infty$, the Ito integral $\int_0^t \phi(s) db(s)$ has the properties

$$E \int_0^t \phi(s) db(s) = 0 \quad t < T \quad (2.1)$$

$$E \left(\int_0^t \phi(s) db(s) \right)^2 = E \int_0^t \phi^2(s) ds \quad t < T . \quad (2.2)$$

Let

$$\hat{L}^2([0,T]^r) = \{f \in L^2([0,T]^r) \mid f \text{ is symmetric}\} .$$

This will be the set of integrands for the rth order integral. If $f \in L^2([0,T]^r)$,

$f(\sigma, \dots) \in L^2([0,T]^{r-1})$, will denote the section of f at σ .

Definition 2.1 Let $f \in L^2([0,T]^r)$ $t < T$. $I_t^r(f)$ is defined recursively by $(L^2([0,T]^0))$
 $: = \mathbb{R}$)

$$I_t^0(f) = f \quad \text{for } r = 0 \quad (2.3)$$

$$I_t^r(f) = \int_0^t I_s^{r-1}(f(s, \dots)) db(s) \quad r > 0 .$$

$I_t^r(f)$ is the rth order multiple integral of f with respect to $b(\cdot)$ up to time t . Alternately stated,

$$I_t^r(f) = \int_0^t \int_0^{s_1} \dots \int_0^{s_{r-1}} f(s_1, \dots, s_r) db(s_r) \dots db(s_1) .$$

To insure that the right-hand side of (2.3) is well defined it suffices to show that

$I_s^{r-1}(f(s, \dots))$ has a jointly measurable version with bounded $L^2(\Omega \times [0,T], P \times \lambda)$ norm.

This may be done by proving recursively, along with the definition, that

$$\begin{aligned} EI_t^r(f) I_t^r(g) &= \frac{1}{r!} (f, g) \\ &= \int_0^t \int_0^{s_1} \dots \int_0^{s_{r-1}} f(s_1, \dots, s_r) g(s_1, \dots, s_r) ds_r \dots ds_1 \end{aligned} \quad (2.4)$$

for all $f, g \in L^2([0,T]^r)$. This is a consequence of (2.2). Then, if f^n is a sequence of symmetrizations of separable functions, such that $f^n \rightarrow f$ in L^2 -norm, $I_s^r(f^n(s, \dots))$ is jointly measurable for all n and

$$\lim_{n \rightarrow \infty} E \int_0^T [I_s^{r-1}(f^n(s, \dots)) - I_s^{r-1}(f(s, \dots))]^2 ds = 0 .$$

Thus we can find a jointly measurable version of $I_s^{r-1}(f(s, \dots))$.

It is important to note that multiple integrals have zero mean and that integrals of different orders are orthogonal; that is, for $f \in L^2([0, T]^r)$, $g \in L^2([0, T]^q)$, $q \neq r$, t , $s < t$

$$E I_t^r(f) = 0 \quad (2.5)$$

$$E[I_t^r(f) I_s^q(g)] = 0 .$$

These follow from repeated application of (2.1) and (2.2).

Remark. The requirement of symmetry for the integrands is not necessary, since integration is carried out only over the set where $s_1 > s_2 > \dots > s_r$. However this convention is convenient in formulating the multiplication formula in section 2.2.

The following technical lemma, a Fubini result on interchanging db and ds integrations, is needed later.

Lemma 2.1 Let $f \in L^2([0, T]^r)$. For $t < T$

$$\int_0^t I_s^{r-1}(f(s, \dots)) ds = \int_0^t \dots \int_{s_1}^{s_{r-2}} \int_{s_1}^t f(u, s_1, \dots, s_{r-1}) du db(s_{r-1}) \dots db(s_1) . \quad (2.6)$$

Proof Define $g_t(s_1, \dots, s_{r-1}) = \int_{s_1}^t f(u, s_1, \dots, s_{r-1}) du$. The r.h.s. of (2.6) is $I_t^{r-1}(g_t)$. To prove the lemma, simply verify that

$$E \left[\int_0^t I_s^{r-1}(f(s, \dots)) ds - I_t^{r-1}(g_t) \right]^2 = 0$$

by using the basic properties (2.5) of the multiple stochastic integral.

For filtering applications, we must also define multiple integrals

$$\int_0^t \dots \int_0^{s_{r-1}} f(s_1, \dots, s_r) dy(s_r) \dots (dy(s_1)) \quad (2.7)$$

with respect to observation semi-martingales

$$y(t) = \int_0^t h(x_s) ds + w(t) \quad (2.8)$$

(the assumptions of section 1 are assumed to be in force). Such integrals are known and have been studied in the context of semi-martingale theory. However, the special structure of (2.8) allows a simple definition which we present here. This takes advantage of the absolute continuity of the $y(\cdot)$ process with respect to Brownian motion; as stated above, if (Ω, \mathcal{F}, P) is the underlying probability space, there exists a probability measure P_0 such that $P_0 \ll P$, $P \ll P_0$, and $y(\cdot)$ is Brownian on $(\Omega, \mathcal{F}, P_0)$. Therefore, for $f \in L^2([0, T]^k)$, we define (2.6) as the random variable, which on $(\Omega, \mathcal{F}, P_0)$ equals the Brownian motion integral defined above. We call this integral $I_t^r(f)$ without reference to measure or process, which should always be clear from context.

The iterative property of $I_t^r(f)$ remains true for dy integrals; that is,

$$I_t^r(f) = \int_0^t I_s^{r-1}(f(s, \dots)) dy(s), \quad (2.9)$$

where the integral in (2.9) is defined with respect to the semi-martingale $y(\cdot)$ in the usual sense (see Liptser and Shirayev [11]). However, neither the expression (2.4) nor the orthogonality of different orders, (2.5), now holds. Instead, we can prove the following lemma, which is useful in section 3.2. (In this discussion, we abbreviate $h(s, x(s))$ by $h(s)$.)

Lemma 2.2 Suppose $E[\int_0^T h^2(s) ds]^r < \infty$. Then for $k < r$ and $f \in L^2([0, T]^k)$

(i) $E[I_t^k(f)]^2 \leq M_k \|f\|_{L^2}^2$; $M_k < \infty$ is independent of f

(ii) $E I_t^k(f) = \int_0^t \cdots \int_0^{s_{k-1}} f(s_1, \dots, s_k) E[h(s_1) \cdots h(s_k)] ds_k \cdots ds_1$.

Proof. We will actually prove by induction the more general result: for $r > l > k$

$$\sigma_k, \dots, \sigma_l \in [0, T]$$

$$E[x(\sigma_l) \dots x(\sigma_{k+1}) I_{\sigma_k}^k(f)]^2 \leq a_{l,k}(\sigma_{k+1}, \dots, \sigma_l) \|f\|^2 \quad (2.10)$$

where $a_{l,k} \in L^1([0, T]^{l-k})$, and

$$E[h(\sigma_l) \dots h(\sigma_{k+1}) I_{\sigma_k}^k(f)] = \int_0^{\sigma_k} \int_0^{s_1} \dots \int_0^{s_{k-1}} f(s_1, \dots, s_k) E[h(s_1) \dots h(s_k) h(\sigma_{k+1}) \dots h(\sigma_l)] ds_k \dots ds_1. \quad (2.11)$$

Lemma 2.2 is the case $l = k$ for every $k < r$. First we demonstrate (2.10) and (2.11) for $r > l > k = 1$, using the iterative formula of (2.9) and the independence of $x(\cdot)$ and $w(\cdot)$. Thus

$$\begin{aligned} E[h(\sigma_l) \dots h(\sigma_2) \int_0^{\sigma_1} f(s) dy(s)]^2 &= \\ E[h(\sigma_l) \dots h(\sigma_2) \{ \int_0^{\sigma_1} f(s) h(s) ds + \int_0^{\sigma_1} f(s) dw(s) \}]^2 &\leq \\ (2E \int_0^T [h(\sigma_l) \dots h(s)]^2 ds + 2 E[h(\sigma_l)^2 \dots h(\sigma_2)^2]) \|f\|^2 &= a_{l,1}(\sigma_2, \dots, \sigma_l) \|f\|^2. \end{aligned} \quad (2.12)$$

To derive the inequality in (2.12), the Cauchy-Schwarz inequality is used several times.

$a_{l,1} \in L^1([0, T]^{l-1})$ for $l < r$ because $E[\int_0^T h^2(s) ds]^r < \infty$. Likewise

$$\begin{aligned} E[h(\sigma_l) \dots h(\sigma_2) \int_0^{\sigma_1} f(s) dy_s] &= \\ E[h(\sigma_l) \dots h(\sigma_2) \{ \int_0^{\sigma_1} f(s) h(s) ds + \int_0^{\sigma_1} f(s) dw(s) \}] &= \int_0^{\sigma_1} f(s) E[h(\sigma_2) \dots h(\sigma_l)] ds. \end{aligned} \quad (2.13)$$

Now suppose (2.10) and (2.11) are true for a fixed k and all $\ell, r > \ell > k$. Again, using $I_t^{k+1}(f) = \int_0^t I_s^k(f(s, \dots)) dy(s)$, Cauchy-Schwarz, and induction

$$\begin{aligned} E[h(\sigma_\ell) \cdots h(\sigma_{k+2}) I_{\sigma_{k+1}}^{k+1}(f)]^2 &\leq \\ 2 \int_0^{\sigma_{k+1}} \int_0^{\sigma_{k+1}} E[h(\sigma_\ell) \cdots h(\sigma_{k+2}) h(s_1) I_{s_2}^k(f(s_2, \dots))]^2 ds_2 ds_1 \\ + 2 \int_0^{\sigma_k} E[h(\sigma_\ell) \cdots h(\sigma_{k+2}) I_s^k(f(s, \dots))]^2 ds \\ &\leq [2 \int_0^T a_{\ell, k}(s, \sigma_{k+2}, \dots, \sigma_\ell) ds + 2a_{\ell-1, k}(\sigma_{k+2}, \dots, \sigma_\ell)] \|f\|^2 \\ &= a_{\ell, k+1}(\sigma_{k+2}, \dots, \sigma_\ell) \|f\|^2 . \end{aligned}$$

By induction, $a_{\ell, k+1} \in L^1([0, T]^{\ell-k-1})$. Thus (2.10) is true for $k+1$. That (2.10) holds for k also implies

$$E \int_0^T I_s^k(f(s, \dots)) ds < \infty .$$

Thus, from (2.1),

$$E \int_0^t I_s^k(f(s, \dots)) dw(s) = 0, \text{ for } t \leq T .$$

With the aid of this equality we can prove that (2.11) also is true for $k+1$. This completes the induction step. Induction stops at $k=r$ since we have required $r > \ell > k$ in order to apply $E(\int_0^T h^2(s) ds)^r < \infty$.

2.3 The multiplication formula.

As above let $(b(t), F_t^b)$ denote a standard Brownian motion. If $\psi(b(s), s < t)$ is a functional of $b(\cdot)$ up to time t , we want to consider expansions of the form

$$\psi = \sum_{r=0}^{\infty} I_t^r (k^r) .$$

(If $\psi \in L^2(\Omega, F_t^b, P)$ such a representation exists, uniquely, and the series converges to ψ in mean-square; see Ito [7] or Hida [5].) Rules prescribing how this representation changes as various operations are performed on ψ must be available if multiple integral expansions are to be of use in applications. In this section, we address the simplest problem in this direction. If $f \in \hat{L}^2([0, T]^r)$ and $g \in \hat{L}^2([0, T]^q)$, what, if any, are the kernels, $\{I_t^i\}_{i=0}^{\infty}$, such that

$$I_t^r(f) I_t^q(g) = \sum_{i=0}^{\infty} I_t^i (I_t^i) \quad (t < T) ?$$

To express the answer, we first introduce the following notation.

Definition 2.2

(i) $P_r \equiv$ projection of $L^2([0, T]^r)$ onto $\hat{L}^2([0, T]^r)$:

$$(P_r h)(\sigma_1, \dots, \sigma_r) = \frac{1}{r!} \sum_{\pi \in S_r} h(\sigma_{\pi(1)}, \dots, \sigma_{\pi(r)})$$

where S_r = permutation group on r letters.

(ii) For integers r, q, k , $0 \leq k \leq \min(r, q)$, and

$$f \in \hat{L}^2([0, T]^r), g \in \hat{L}^2([0, T]^q)$$

$$(f \otimes_k (t) g)(\sigma_1, \dots, \sigma_{r+q-2k})$$

$$\equiv \frac{1}{k!} \int_0^t \dots \int_0^t f(s_1, \dots, s_k, \sigma_1, \dots, \sigma_{r-k}) g(s_1, \dots, s_k, \sigma_{r-k+1}, \dots, \sigma_{r+q-2k}) ds_k \dots ds_1$$

$$(iii) f \otimes_k (t) g \equiv P_{r+q-2k}[f \otimes_k (t) g]$$

$$(iv) f \otimes g = f \otimes_0(t)g = P_{r+q}[f(\sigma_1, \dots, \sigma_r)g(\sigma_{r+1}, \dots, \sigma_{r+q})]$$

$\otimes_k(t)$ is the operation by which new kernels are created from old; indeed,

$$f \otimes_k(t)g : L^2([0, T]^r) \times L^2([0, T]^q) \rightarrow L^2([0, T]^{r+q-2k})$$

as the following lemma demonstrates

Lemma 2.3 For every $t < T$

$$f \otimes_k(t)g \in L^2([0, T]^{r+q-2k})$$

In fact,

$$\|f \otimes_k(t)g\|^2 \leq c_{r,q,k} \|f\|^2 \|g\|^2$$

where $c_{r,q,k}$ is independent of f and g .

Proof It suffices to prove the lemma for \otimes , instead of \otimes_0 , since P_{r+q-2k} is a bounded operator. Let $d\bar{\sigma} = d\sigma_1 \cdots d\sigma_{r+q-2k}$, $d\bar{s} = ds_1 \cdots ds_k$. We then have, using the Cauchy-Schwarz inequality

$$\begin{aligned} \|f \otimes_k(t)g\| &= \left(\int_{[0, T]^{r+q-2k}} \frac{d\bar{\sigma}}{(k!)^2} \times \left[\int_{[0, T]^k} d\bar{s} f(s_1, \dots, s_k, \sigma_1, \dots) g(s_1, \dots, s_k, \dots, \sigma_{r+q-2k}) \right]^2 \right)^{1/2} \\ &\leq \frac{1}{(k!)^2} \|f\|^2 \|g\|^2. \end{aligned}$$

To understand the meaning of $\otimes_k(t)$, it is useful to think of the functions f and g as tensors, which they in fact are by the isomorphism

$$L^2([0, T]^r) \cong L^2([0, T]) \otimes \cdots \otimes L^2([0, T]) \quad (\text{r-fold}).$$

Then $f \otimes_k(t)g$ may be viewed as a tensor contraction since it 'sums', that is, integrates, f and g along the first k indices. Thus $f \otimes_k(t)g$ is simply a symmetrized, k -fold, tensor contraction. It is in this definition that the symmetry of f and g is used; otherwise $\otimes(t)$ would have a more complicated definition. For notational convenience, we shall often write \otimes_k instead of $\otimes_k(t)$, in which case the (t) is to be assumed. When the time parameter is important or different than t , it will always be given.

We can now state the result

Theorem 2.1 Let $f \in \hat{L}^2([0, T]^r)$, $g \in \hat{L}^2([0, T]^q)$. Then

$$I_t^r(f) I_t^q(g) = \sum_{k=0}^{\min(r,q)} I_t^{r+q-2k} ((\frac{r+q-2k}{r-k}) f_k \otimes (t) g_k) . \quad (2.14)$$

- Remarks
1. (2.14) shall be referred to as the multiplication formula. Hida has independently derived this result as an application of his theory of generalized Brownian functionals (personal communication of T. Hida; for generalized Brownian functional theory, see Hida [4]). Our proof is elementary, using only Ito's differentiation rule. For similar theory, see also Meyer [15]. Versions of this formula are also known in mathematical quantum field theory (Reed, Simon [19]). See Mitter and Ocone [17] for further comments.
 2. The multiplication formula generalizes a Hermite polynomial identity. The nth order Hermite polynomial of a single variable is

$$h_n(x) = \frac{(-1)^n}{\sqrt{n}} e^{-x^2/2} \frac{d^n}{dx^n} e^{-x^2/2} .$$

Let

$$\bar{G}_r = \{I_t^r(f) \mid f \in \hat{L}^2([0, t]^r)\}$$

and let $\{\phi_n\}_{n=1}^\infty$ be an orthonormal basis of $L^2([0, t])$. Then, (Ito [7], Kallianpur [8])

$$\bar{G}_r = \overline{\text{Sp}} \left\{ \prod_{i=1}^n h_{p_i} \left(\int_0^t \phi_{j_i}(s) dB(s) \right) \mid \begin{array}{l} p_1 + \dots + p_n = r \\ j_1, \dots, j_n \text{ are pairwise unequal} \end{array} \right\} ,$$

where $\overline{\text{Sp}}$ denotes the closure in $L^2(\mathbb{P})$ of the linear span. One then sees that (2.14) generalizes the identity, ([12]),

$$h_r(x) h_q(x) = \sum_{k=0}^{\min(r,q)} \left(\binom{r}{k} \binom{q}{k} \binom{r+q-2k}{r-k} \right)^{1/2} h_{r+q-2k}(x) \quad (2.15)$$

$\forall r, q > 0$. There is a discrepancy between (2.15) and (2.14) in the factors multiplying the expansion terms, but this is due to the different normalizations involved in the definitions of h_n , I^r and O . The relationship between (2.14) and (2.15) may be seen clearly in Hida's work, but we shall not pursue the matter further here.

We will show how to prove theorem 2.1 using Ito's rule and induction. For this purpose, we need certain facts and identities concerning O , and these are collected in the next lemma. The notation $f(s_1, \dots, s_k, \dots)$ indicates the section of f in which the first k variables are fixed at s_1, \dots, s_k respectively.

Lemma 2.4

$$(i) f(\sigma_1, \dots, O_k(\sigma_1)g(\sigma_1, \dots) (\sigma_2, \dots, \sigma_{r+q-2k-1}) \in L^2([0, T]^{r+q-2k})$$

$$(ii) f O_k(t)g = f O_k(\sigma)g + \int_0^t f(s, \dots) O_{k-1}(s) g(s, \dots) ds \quad (2.16)$$

$$\begin{aligned} (iii) \text{ For } k > 1, & (f O_k(t)g)(\sigma_1, \dots, \sigma_{r+q-2k}) \\ &= [\frac{r-k}{r+q-2k} f(\sigma_1, \dots, O_k(t)g) + \frac{q-k}{r+q-2k} f O_k(t)g(\sigma_1, \dots)] (\sigma_2, \dots, \sigma_{r+q-2k}) \end{aligned} \quad (2.17)$$

$$\begin{aligned} (iv) & (f O(t)g)(\sigma_1, \dots, \sigma_{r+q}) = \\ & [\frac{r}{r+q} f(\sigma_1, \dots, O(t)g) + \frac{q}{r+q} f O(t)g(\sigma_1, \dots)] (\sigma_2, \dots, \sigma_{r+q-2k}) . \end{aligned} \quad (2.18)$$

Proof i) follows from calculations similar to those in lemma 2.3. The details will not be presented.

ii) By direct calculation and definition, using the symmetry of f and g extensively.

$$\begin{aligned} f O_k(t)g &= P_{r+q-2k} \left[\frac{1}{k!} \int_0^t \dots \int_0^t ds_1 \dots ds_k f(s_1, \dots, s_k, \dots) g(s_1, \dots, s_k, \dots) \right] \\ &= P_{r+q-2k} \left(\int_0^t \int_0^{s_1} \dots \int_0^{s_{k-1}} f(s_1, \dots, s_k, \dots) g(s_1, \dots, s_k, \dots) ds_k \dots ds_1 \right) \\ &= P_{r+q-2k} \left(\frac{1}{k!} \int_0^\sigma \int_0^\sigma \dots \int_0^\sigma f(s_1, \dots, s_k, \dots) g(s_1, \dots, s_k, \dots) ds_k \dots ds_1 \right) \end{aligned}$$

$$+ p_{r+q-2k} \frac{1}{(k-1)!} \int_0^t ds \int_0^{s_1} \dots \int_0^{s_{k-1}} f(s_1, \dots, s_k, \dots) g(s_1, \dots, s_k, \dots) ds_k \dots ds_1 \\ = f \otimes_k (s) g + \int_0^t ds f(s, \dots) \otimes_{k-1} (s) g(s, \dots)$$

(iii) and (iv). The proofs of (iii) and (iv) are similar, (iv) being just a special case of (iii). We shall only present (iv), as it is simpler. Note first that, by definition,

$$\frac{r}{r+q} [f(\sigma_1, \dots) \otimes (t) g](\sigma_2, \dots, \sigma_{r+q}) \\ = \frac{r}{r+q} \frac{1}{(r+q-1)!} \sum_{\pi \in S_{r+q-1}} f(\sigma_1, \sigma_{\pi(2)}, \dots, \sigma_{\pi(r)}) g(\sigma_{\pi(r+1)}, \dots, \sigma_{\pi(r+q)}) \quad (2.19)$$

where $\pi \in S_{r+q-1}$ is interpreted as a permutation of $\{2, \dots, r+q\}$. Now using the symmetry of f , (2.19) may be written as:

$$\frac{1}{(r+q)!} \sum_{j=1}^r \sum_{\pi \in S_{r+q-1}} f(\sigma_{\pi(2)}, \dots, \sigma_{\pi(j)}, \sigma_1, \sigma_{\pi(j+1)}, \dots, \sigma_{\pi(r)}) \times \\ g(\sigma_{\pi(r+1)}, \dots, \sigma_{\pi(r+q)}) \quad (2.20)$$

Using the expression analogous to (2.20) for $\frac{q}{r+q} f \otimes (t) g(\sigma_1, \dots)$,

$$\left(\frac{r}{r+q} f(\sigma_1, \dots) \otimes (t) g + \frac{q}{r+q} f \otimes (t) g(\sigma_1, \dots) \right) (\sigma_2, \dots, \sigma_{r+q}) \\ = \frac{1}{(r+q)!} \left\{ \sum_{j=1}^r \sum_{\pi \in S_{r+q-1}} f(\sigma_{\pi(2)}, \dots, \sigma_1, \dots, \sigma_{\pi(r)}) \times g(\sigma_{\pi(r+1)}, \dots, \sigma_{\pi(r+q)}) \right. \\ \left. + \sum_{j=1}^q \sum_{\pi \in S_{r+q-1}} f(\sigma_{\pi(2)}, \dots, \sigma_{\pi(r+1)}) \times g(\sigma_{\pi(r+2)}, \dots, \sigma_1, \dots, \sigma_{\pi(r+q)}) \right\} \\ = \frac{1}{(r+q)!} \sum_{\pi \in S_{r+q}} f(\sigma_{\pi(1)}, \dots) g(\sigma_{\pi(r+1)}, \dots, \sigma_{\pi(r+q)}) \\ = f \otimes (t) g(\sigma_1, \dots, \sigma_{r+q}) \quad .$$

This is the desired result.

Proof of theorem 2.1. We use Ito's differentiation formula and the preceding lemmas to implement an induction argument that proceeds in two steps:

- (a) Show (by induction) that (2.14) holds for orders $r = n$, $q = 1$, $\forall n$.
- (b) Assuming (2.14) for $(r-1,q)$, $(r,q-1)$ and $(r-1,q-1)$, show that it holds for (r,q) .
(a) and (b) then provide a consistent scheme of induction for proving theorem 2.1 for all orders.

Step (a) By Ito's differentiation rule

$$\int_0^t f(s)db(s) \int_0^t g(s)db(s) = \int_0^t \int_0^{s_1} [f(s_1)g(s_2) + f(s_2)g(s_1)]db(s_1)$$
$$+ \int_0^t f(s)g(s)ds .$$

This proves the case $r = q = 1$.

Suppose that the theorem is true for $(r,q) = (n-1,1)$ and let $f \in \hat{L}^2([0,T]^n)$, $g \in \hat{L}^2([0,T])$. Applying Ito's differentiation rule again,

$$I_t^n(f)I_t^1(g) = \int_0^t g(s) I_s^n(f)db(s) + \int_0^t I_s^{n-1}(f(s,\dots))I_s^1(g)db(s)$$
$$+ \int_0^t I_s^{n-1}(g(s)f(s,\dots))ds . \quad (2.21)$$

By induction,

$$I_s^{n-1}(f(s,\dots))I_s^1(g) = I_s^n(n[f(s,\dots) \circ g]) + I_s^{n-2}(f(s,\dots) \circ_1(s)g) .$$

Lemma 2.5(i) and lemma 2.1 justify interchanging integrations in the last term of (2.21):

$$\int_0^t I_s^{n-1}(g(s)f(s,\dots))ds = I_t^{n-1}(\int_{s_1}^t g(u)f(u,s_1,\dots,s_{n-1})du) .$$

Thus, by substitution in (2.20)

$$\begin{aligned}
I_t^n(f) I_t^1(g) &= \int_0^t [I_s^n(g(s)f(\cdot)) + I_s^n(n[f(s,\cdot) \otimes g])] db(s) \\
&\quad + \int_0^t I_s^{n-2}(f(s,\cdot) \otimes_1(s) g) db(s) \\
&\quad + I_t^{n-1}(\int_{s_1}^t g(u)f(u,s_1,\dots,s_{n-1}) du) \\
&= I_k^{n+1}\{g(\sigma_1)f(\sigma_2,\dots,\sigma_n) + n[f(\sigma_1,\dots) \otimes g(\sigma_1)](\sigma_2,\dots,\sigma_n)\} \\
&\quad + I_t^{n-1}\{f(\sigma_1,\dots) \otimes_1(\sigma_1) g(\sigma_2,\dots,\sigma_n) + \int_0^t g(s)f(s,\sigma_1,\dots,\sigma_{n-1}) ds\}.
\end{aligned}$$

By lemma 2.5 (iii) and (iv) this becomes

$$I_t^{n+1}((n+1)f \otimes g) + I_t^{n-1}(f \otimes_1(t) g),$$

which completes the induction step of (a).

Step b Without loss of generality assume that $q < r$. The induction hypothesis is that theorem 2.4 is true for $(r-1,q)$, $(r,q-1)$, and $(r-1,q-1)$. Apply Ito's differentiation rule:

$$\begin{aligned}
I_t^r(f) I_t^q(g) &= \int_0^t I_s^q(g) I_s^{r-1}(f(s,\cdot)) db(s) \\
&\quad + \int_0^t I_s^{q-1}(g) I_s^r(f) db(s) \\
&\quad + \int_0^t I_s^{r-1}(f(s,\cdot)) I_s^{q-1}(g(s,\cdot)) ds.
\end{aligned}$$

Next, use the induction hypothesis to expand the integrands in (2.22), then interchange ds and $db(s)$ integrations where necessary, and collect like order terms. The result is, for $q < r$

$$\begin{aligned}
I_t^r(f) I_t^q(g) &= I_t^{r+q}\{[(\frac{r+q-1}{r-1})(f(s_1,\dots) \otimes g) + I_r^{r+q-1}(f \otimes g(s_1,\dots))](s_2,\dots,s_{r+q})\} \\
&\quad + \sum_{k=1}^{q-1} I_t^{r+q-2k}(\frac{r+q-1-2k}{r-1-k})(f(s_1,\dots) \otimes_k(s_1,\dots))(s_2,\dots,s_{r+q-2k}) \\
&\quad + (\frac{r+q-1-2k}{r-k})(f \otimes_k(s_1) g(s_1,\dots))(s_2,\dots,s_{r+q-2k})
\end{aligned}$$

$$\begin{aligned}
& + \left(\frac{r+q-2k}{r-k} \right) \int_{s_1}^t f(u, \dots) \theta_{k-1}(u) g(u, \dots) du \\
& + I_t^{r-q} \{ (f(s_1, \dots) \theta_q(s_1) g(s_2, \dots, s_{r-q}) + \int_{s_1}^t f(u, \dots) \theta_{q-1}(u) g(u, \dots) du) .
\end{aligned}$$

To complete the proof, we need only apply the identities of lemma 2.4 (iii) and (iv) to the kernels of this last expression. For example, the kernel of I_t^{r+q-2k} , $1 \leq k \leq q-1$, equals

$$\left(\frac{r+q-2k}{r-k} \right) \left[\frac{r-k}{r+q-2k} (f(s_1, \dots) \theta_k(s_1) g(s_2, \dots) + \frac{q-k}{r+q-2k} [f \theta_k(s_1) g(s_1, \dots)] (s_2, \dots)$$

$$\begin{aligned}
& + (\int_{s_1}^t f(u, \dots) \theta_k(u) g(u, \dots) du) (s_2, \dots)) \\
& = \left(\frac{r+q-2k}{r-k} \right) [(f \theta_k(s_1) g)(s_1, \dots, s_{r+q-2k}) + (\int_{s_1}^t f(u, \dots) \theta_{k-1}(u) g(u, \dots) du) (s_2, \dots) \\
& = \left(\frac{r+q-2k}{r-k} \right) [(f \theta_k(t) g)(s_1, \dots, s_{r+q-2k})] .
\end{aligned}$$

This is the kernel given in (2.14). The kernels of I_t^{r+q-2k} , $k = 0$ and $k = q$ are treated similarly. This completes the proof. ■

3. Multiple Integral Expansions in Filtering Theory.

This section explores the use of multiple integral expansions for optimal and suboptimal filtering. The estimation problem considered is the general problem stated in the introduction, and the notations and assumptions established there shall remain in force. For additional notational convenience, we let $f(t) := f(t; x_s, s < t)$, $h(s) := h(s, x(s))$ and $f_t := E[f(t) | F_t^Y]$.

3.1. Expansion of the optimal filter

In theorem 3.1 below we derive an expression for \hat{f}_t as a ratio of two multiple integral expansions in which the process of integration is $y(t)$, the observation semi-martingale, and the integrands are deterministic functionals computable from the

(unconditioned) distribution of the signal process. First we state some preliminary definitions and a lemma.

Let

$$L_t := \exp\left(\int_0^t h(s)dy(s) - \frac{1}{2}\int_0^t h^2(s)ds\right) .$$

L_t is the important process in this calculation. Observe that $L_t = \frac{dP}{dP_0}$ and $(L_t, F_t^X \vee F_t^Y)$ is a martingale on $(\Omega, \mathcal{F}, P_0)$, ($F_t^X := \sigma(x(s); s \in \mathbb{R}^+)$). A conditioning argument then shows that the Kallianpur-Striebel formula, (1.2), can be expressed as

$$f_t = \frac{\mathbb{E}_0\{f(t)L_t \mid F_t^Y\}}{\mathbb{E}_0\{L_t \mid F_t^Y\}} . \quad (3.1)$$

The following process, based on L_t , will also appear:

$$L_t^{(r)} := \int_0^t \dots \int_0^{s_r} h(s_1) \dots h(s_r) dy(s_r) \dots dy(s_1) .$$

Note that $L_t^{(r)}$ is not a multiple integral of the type defined in §2 since the integrand is not deterministic. $L_t^{(r)}$ may be properly defined by noticing that

$$L_t^{(r)} = \int_0^t h(s) L_s^{(r-1)} dy(s) .$$

Iterative use of the stochastic Ito integral then specifies $L_t^{(r)}$ for any order r . This is especially easy to carry out on $(\Omega, \mathcal{F}, P_0)$, on which $y(t)$ is a Brownian motion independent of the signal (see theorem 1.1).

The following stochastic Fubini theorem for interchanging conditional expectation and stochastic integration is needed; it is a direct consequence of theorem 5.14 in Liptser and Shirayev [11].

Lemma 3.1 Let $\phi(s)$ be a $F_t^X \vee F_t^Y$ adapted process such that

$$E_0[\int_0^T \phi^2(s) ds] < \infty$$

$$\text{Then } E\{\int_0^T \phi(s) dy(s) \mid F_t^Y\} = \int_0^t E_0\{\phi(s) \mid F_s^Y\} dy(s).$$

Finally, it is convenient to introduce the functions

$$I_n(t, s_1, \dots, s_n) := E\{f(t)h(s_1)\dots h(s_n)\} \quad n \geq 0$$

$$k_n(s_1, \dots, s_n) := E\{h(s_1)\dots h(s_n)\} \quad n \geq 1$$

$$k_0 := 1.$$

Theorem 3.1

i) (Partial expansion) If $E(\int_0^t h^2(s) ds)^x < \infty$, and $E[f^2(t)(\int_0^t h^2(s) ds)^x] < \infty$,

then

$$\hat{f}_t = \frac{\sum_{n=0}^x I_t^{(n)}(I_n(t)) + E_0\{f(t) L_t^{(x)} \mid F_t^Y\}}{\sum_{n=0}^x I_t^{(n)}(k_n) + E_0\{f(t) L_t^{(x)} \mid F_t^Y\}}. \quad (3.2)$$

ii) (Full expansion) If $E(\exp \int_0^t h^2(s) ds) < \infty$, and $E[f^2(t) \exp \int_0^t h^2(s) ds] < \infty$

$$\hat{f}_t = \frac{\sum_{n=0}^{\infty} I_t^{(n)}(I_n(t))}{\sum_{n=0}^{\infty} I_t^{(n)}(k_n)} \quad (3.3)$$

and the expansions converge in $L^1(P)$.

Proof:

Part i) By applying Ito's differentiation rule to L_t ,

$$dL_s = h(s)L_s du(s)$$

so that

$$L_t = 1 + \int_0^t h(s)L_s dy(s). \quad (3.4)$$

Iterating (4.4), we find that for any r

$$L_t = 1 + \int_0^t h(s)dy(s) + \int_0^t \int_0^{s_1} h(s_1)h(s_2)dy(s_2)dy(s_1) \\ + \dots + L_t^{(r)} .$$

Now substitute this expansion into the Kallianpur-Striebel formula 3.1 for \hat{f}_t . The denominator, for example, becomes

$$E_0\{L_t | F_t^Y\} = 1 + \sum_{n=1}^r E_0\{\int_0^t \dots \int_0^{s_{n-1}} h(s_1) \dots h(s_n) \dots dy(s_1) | F_t^Y\} \\ + E_0\{L_t^{(r)} | F_t^Y\} . \quad (3.5)$$

The hypothesis $E[\int_0^t h^2(s)ds]^r < \infty$ of (i) allows lemma 3.1 to be applied to the terms of (3.5), with the result,

$$E_0\{L_t | F_t^Y\} = 1 + \sum_{n=1}^r \int_0^t \dots \int_0^{s_{n-1}} E_0\{h(s_1) \dots h(s_n)\} dy(s_n) \dots dy(s_1) + E\{L_t^{(r)} | F_t^Y\} .$$

Since the distribution of the signal process is invariant under the change of measures from P to P_0

$$E_0\{h(s_1) \dots h(s_n)\} = E\{h(s_1) \dots h(s_n)\} \\ = k_n(s_1, \dots, s_n) .$$

Therefore

$$E_0\{L_t | F_t^Y\} = \sum_{n=0}^r I_t^{(n)}(k_n) + E\{L_t^{(r)} | F_t^Y\} .$$

A similar calculation yields

$$E_0\{f(t)L_t | F_t^Y\} = \sum_{n=0}^r I_t^{(n)}(l_n(t)) + E\{f(t)L_t^{(r)} | F_t^Y\} .$$

Substitution of these expressions into the Kallianpur-Striebel formula then proves (3.2). Part ii). Formally, the proof of the full expansion follows by setting $r = \infty$. To prove it rigorously, we first show that $E \exp[\int_0^T h^2(s)ds] < \infty$ implies

$$L_t = m.s.(P_0) \lim_{N \rightarrow \infty} [1 + \sum_{n=1}^N \int_0^t \cdots \int_0^{s_{n-1}} h(s_1) \cdots h(s_n) dy(s_n) \cdots dy(s_1)] . \quad (3.6)$$

Denote the finite series on the right hand side of (3.6) by A_t^N . Then

$$E_0 (L_t - A_t^N)^2 = E_0 [\int_0^t \cdots \int_0^{s_N} h(s_1) \cdots h(s_{N+1}) L_{s_{N+1}} dy(s_{N+1}) \cdots dy(s_1)]^2 .$$

By employing the standard computational rules (2.1), (2.2) for stochastic integrals, this last expression equals

$$\int_0^t \cdots \int_0^{s_N} E_0 [h^2(s_1) \cdots h^2(s_N) L_{s_{N+1}}^2] ds_{N+1} \cdots ds$$

provided that it is finite. However,

$$\begin{aligned} & E_0 [h^2(s_1) \cdots h^2(s_N) L_{s_{N+1}}^2] \\ &= E_0 [h^2(s_1) \cdots h^2(s_N) \exp[-\int_0^{s_{N+1}} h^2(s)ds] E_0 [\exp[2 \int_0^{s_{N+1}} h(s)dy(s)] | F_{s_1}^X]] . \quad (3.7) \end{aligned}$$

Now on (Ω, P_0) , $x(\cdot)$ and $y(\cdot)$ are independent and $y(\cdot)$ is Brownian, and hence, given

$\{x(s), s < s_1\}$, $\int_0^{s_{N+1}} h(s)dy(s)$ is a Gaussian random variable with mean 0 and variance $\int_0^{s_{N+1}} h^2(s)ds$. Thus

$$E_0 [\exp 2 \int_0^{s_{N+1}} h(s)dy(s) | F_{s_1}^X] = \exp 2 \int_0^{s_{N+1}} h^2(s)ds . \quad (3.8)$$

Therefore, using (3.8) in (3.7)

$$(3.7) = E_0 \{ h^2(s_1) \dots h^2(s_{N+1}) \exp \{ \int_0^{s_{N+1}} h^2(s) ds \} \}$$

$$= E_0 \{ h^2(s_1) \dots h^2(s_{N+1}) \sum_0^{\infty} \int_0^{s_{N+1}} \int_0^{s_1} \dots \int_0^{s_{j-1}} h^2(s_1) \dots h^2(s_j) ds_j \dots ds_1 \} .$$

As a result

$$\int_0^t \dots \int_0^{s_{N+1}} E_0 \{ h^2(s_1) \dots h^2(s_{N+1}) L_{s_{N+1}}^2 \} ds_{N+1} \dots s_1$$

$$= \sum_{j=N+1}^{\infty} \int_0^t \dots \int_0^{s_{j-1}} E_0 \{ h^2(s_1) \dots h^2(s_j) \} ds_j \dots ds_1$$

$$= \sum_{j=N+1}^{\infty} \frac{1}{j!} E_0 \{ \int_0^T \dots \int_0^T h^2(s_1) \dots h^2(s_j) ds_j \dots ds_1 \} . \quad (3.9)$$

Since $E \exp \{ \int_0^T h^2(s) ds \} < \infty$, (3.9) tends to 0 as $N \rightarrow \infty$, proving that $L_t =$
m.s. (P_0) $\lim_{N \rightarrow \infty} A_t^N$ for all $t < T$, as desired. Lemma 4.1 can now be invoked for every

order n , so that

$$E_0 \{ L_t | F_t^Y \} = E_0 \{ \text{m.s. } \lim_{N \rightarrow \infty} A_t^N | F_t^Y \}$$

$$= \text{m.s. } \lim_{N \rightarrow \infty} E_0 \{ A_t^N | F_t^Y \}$$

$$= \text{m.s. } (P_0) \lim_{N \rightarrow \infty} [1 + \sum_{n=1}^N I_t^n (k_n)] .$$

A similar proof expands $E_0 \{ f(t) L_t | F_t^Y \}$ in the series

$$I_0(t) + \sum_{n=1}^{\infty} I_t^n (I_n) .$$

Finally, to derive the $L^1(P)$ convergence, note that

$$E_0 \left[\frac{dP}{dP_0} \right]^2 = E_0 L_T^2 = E \left[\exp \int_0^T h^2(s) ds \right] < \infty .$$

Thus

$$\begin{aligned} \left(E_0 [L_t | F_t^Y] - \left(1 + \sum_1^N I_t^n(k_n) \right) \right)^2 \\ < E_0 \left(\frac{dP}{dP_0} \right)^2 E_0 \left(E_0 [L_t | F_t^Y] - \left(1 + \sum_1^N I_t^n(k_n) \right) \right)^2. \end{aligned}$$

Thus, from (3.6), $E_0 [L_t | F_t^Y] = (L^1(P)) \lim_{N \rightarrow \infty} [1 + \sum_{n=1}^N I_t^n(k_n)]$ as claimed. This completes the proof of theorem 3.1.

Let $P(\Delta, t | F_t^Y) = E[\delta_\Delta(x(t)) | F_t^Y]$ denote the conditional distribution of $x(t)$ given the observation up to time t .

Corollary 3.1 If $E[\exp \int_0^T h^2(s) ds] < \infty$

$$p(\Delta, t | F_t^Y) = \frac{E\delta_\Delta(x(t)) + \sum_{n=1}^{\infty} I_t^n(E\delta_\Delta(x(t))h(s_1) \dots h(s_n))}{1 + \sum_{n=1}^{\infty} I_t^n(Eh(s_1) \dots h(s_n))}.$$

A related formula is also of interest. If $x(t)$ has a density $q(x, t)$, $x(t)$ has a conditional density given by

$$p(x, t | F_t^Y) = \frac{E_0 [L(t) | F_t^Y, x(t)=x] q(x, t)}{E_0 [L(t) | F_t^Y]}.$$

Using the same techniques as above, we can easily derive

$$\begin{aligned} E_0 [L(t) | F_t^Y, x(t)=x] q(x, t) &= \left[1 + \sum_{n=0}^{\infty} I_t^n(E[h(s_1) \dots h(s_n) | x(t)=x]) \right] \\ &\times q(x, t) \end{aligned} \tag{3.10}$$

for the numerator of $p(x, t | F_t^Y)$. (3.10) is often called the unnormalized conditional density.

Remark: These results all have an obvious generalization to the multidimensional case.

The Bayes formula (4.1) for \hat{f}_t is properly viewed as the ratio of two conditioned functional integrals, in which the dependencies between $x(\cdot)$ and $y(\cdot)$ are linked in the L_t term. The expansions of theorem 4.1 in effect calculate these functional integrals by expanding L_t . The $x(\cdot)$ and $y(\cdot)$ interactions are then separated in the sense that the

calculation of the filter is decomposed into two parts; first, computation, off-line and prior to filtering, of the kernels λ_n and k_n , and, second, stochastic integration of these kernels against the observations. Of course, in actual practice one can only compute a finite number of terms. In fact, if the kernels are separable or are approximated by separable versions, a truncated expansion may be realized in a finite dimensional and recursive manner, because a stochastic differential system can be constructed to realize any multiple integral with a separable kernel. However, caution must be exercised in approximating the optimal filter by truncations in (4.3), because truncation of the series in the denominator can be a source of severe instability. Although $E\{L(t)|F_t^Y\} > 0$ a.s., a truncation approximation may pass through 0 and so lead to a singularity of the filter. Thus an independent estimate of the denominator is in general required.

Recently, attention has focused on the unnormalized conditional density and the corresponding 'unnormalized' conditional moments, which are just the numerators of the Kallianpur-Striebel formula. E. Wong [22] has given a class of Markov signals for which analytic expressions of (3.10) are available. Again truncation of (4.10) will in general yield functions that attain negative values. For this reason, cumulant expansions

$$p(x,t|F_t^Y) = e^{\sum_0^n I_t^n(g_n(x))}$$

have been studied as an alternate source of approximate filters (see Eterno [1]). We will not pursue these issues further, but instead turn to other theoretical developments based on theorem 3.1.

3.2 Best rth order filters

Finite sums of multiple integrals provide a natural class of causal functionals for the design of suboptimal filters. We introduce the following definition:

Definition 3.1

$$i) \quad Y_r(t) := \{a(t) = \sum_{n=1}^r I_t^n(a_n(t)) | a_n(t, \dots) \in \hat{L}^2([0, t]^n), 0 \leq n \leq r\} .$$

ii) The best rth order estimate of $f(\cdot)$ at time t is an element $\tilde{f}(t) \in Y_r(t)$ such that

$$E(f(t) - \tilde{f}(t))^2 \leq E(f(t) - b(t))^2 \quad (3.11)$$

for all $b(t) \in Y_r(t)$. The kernels of $\tilde{f}(t)$, denoted by $a_0(t), a_1(t), \dots, a_r(t)$, are called the optimal kernels. A process $\tilde{f}(t) \in Y_r(t)$, $t < T$, satisfying (3.11) for $t < T$ is called the best rth order estimate of $f(\cdot)$.

Notice that the best 1st order filter is simply the linear filter, and thus, in the context of multiple integral expansions, best quadratic (2nd order), cubic, quartic, etc. filters are the natural extensions beyond linear filtering.

In this section we characterize the set of optimal kernels as the solution to a system of linear integral equations. The construction of these equations and the proof of their validity utilize the expansion formulae of theorem 3.1 and the multiplication formula for multiple integrals of theorem 2.1. Suppose for the instant that the full expansion (3.3) holds for the optimal filter and that $\tilde{f}(t) = \sum_{n=0}^r I_t^n(a_n(t))$ is an element of $Y_r(t)$, not necessarily the best. If $\tilde{f}(t)$ is to be a good approximation of $\hat{f}(t)$, we want

$$\tilde{f}(t) = \hat{f}(t) = \frac{\sum_{j=0}^r I_t^j(\lambda_j(t))}{\sum_{j=0}^r I_t^j(k_j)}$$

or

$$\tilde{f}(t) \sum_{j=0}^r I_t^j(k_j) = \sum_{j=0}^r I_t^j(\lambda_j(t)) \quad (3.12)$$

Now notice that the left hand side of (3.12) can be rewritten as a multiple integral expansion by applying the multiplication formula. In fact

$$\begin{aligned} \tilde{f}(t) \sum_{j=0}^r I_t^j(k_j) &= \sum_{n=0}^r I_t^n(g_j(t)) \\ g_j(t) &= \sum_{(m,n,i) \in C_j} \binom{m+n-2i}{m-i} a_m(t) o_i(t) k_n \end{aligned} \quad (3.13)$$

where $\mathcal{G}_j := \{(m, n, i) | m+n-2i = j, i < \min(m, n), m < r\}$. Thus, one way to pick an approximation \tilde{f} would be to choose the kernels $a_n(t)$ so that $g_j(t)$ matches $I_j(t)$ for as many orders j as possible. In fact, this is a prescription for the optimal kernels.

Theorem 3.2. Assume $E(\int_0^t h^2(s)ds)^{2r} < \infty$ and $E f^2(t)(\int_0^t h^2(s)ds)^{2r} < \infty$.

Then a best r th order estimate exists. It is given by $\tilde{f}(t) = \sum_{n=0}^r I_t^n(a_n(t))$ iff

$$g_j(t, s_1, \dots, s_j) = E\{f(t)h(s_1)\dots h(s_j)\} \\ (= I_j(t, s_1, \dots, s_j)) \quad (3.14)$$

for $0 < j < r$.

Remark. The equations at (3.14) comprise $r+1$ integral equations for the $r+1$ optimal kernels $a_j(t)$ $0 < j < r$. This can be seen from the definition of $g_j(t)$ and Θ and will be illustrated explicitly in the examples to be discussed.

Before proving theorem 3.2, we first establish some preliminary lemmas. The first deals with existence of estimates.

Lemma 3.2. If $E[\int_0^t h^2(s)ds]^r < \infty$, then the best r th order estimate exists and is unique.

Proof From lemma 2.2 $E[I_t^k(s)]^2 \leq M_k \|a\|_L^2$ for $k < r$. Therefore $\mathbb{Y}_r(t)$ is a mean-square-closed (Hilbert) space of random variables. The lemma follows by the projection theorem.

Of the next two lemmas, the first introduces the optimal estimate to compare suboptimal estimates, and the second verifies a technical identity.

Lemma 3.3. Let $z, v \in L^2(\Omega, \mathcal{F}_t^Y, P)$. Then

$$E(z - f(t))^2 \leq E(v - f(t))^2 \text{ iff } E(z - \hat{f}(t))^2 \leq E(v - \hat{f}(t))^2.$$

Proof. Simply note

$$\begin{aligned} E(z - f(t))^2 &= E(z - \hat{f}(t))^2 + 2E(z - \hat{f}(t))(\hat{f}(t) - f(t)) \\ &\quad + E(\hat{f}(t) - f(t))^2 \\ &= E(z - \hat{f}(t))^2 + E(\hat{f}(t) - f(t))^2. \end{aligned}$$

Lemma 3.4. Let $c(t) + \sum_{n=1}^r I_t^n(c_n(t)) \in Y_r(t)$ and assume that $E(\int_0^T h^2(s)ds)^r < \infty$,

$E\{f^2(t)(\int_0^T h^2(s)ds)^r\} < \infty$, then

$$E_0\{c(t)E_0[L_t^{(r)}|F_t^Y]\} = 0 \quad (3.15)$$

$$E_0\{c(t)E_0\{f(t)L_t^{(r)}|F_t^Y\}\} = 0 \quad (3.16)$$

Proof From (3.5)

$$E_0[L_t^{(r)}|F_t^Y] = E_0[L_t|F_t^Y] - 1 - \sum_{n=1}^r I_t^n(k_n) \quad ,$$

and therefore,

$$\begin{aligned} E_0\{c(t)E_0[L_t^{(r)}|F_t^Y]\} &= E_0\{c(t)E_0[L_t|F_t^Y]\} \\ &\quad - E_0\{c(t)[1 + \sum_{n=1}^r I_t^n(k_n)]\} \quad . \end{aligned} \quad (3.17)$$

Since $y(\cdot)$ is Brownian w.r.t. P_0 ,

$$\begin{aligned} E_0\{c(t)[1 + \sum_{n=1}^r I_t^n(k_n)]\} &= c_0(t) + \\ &\quad \sum_{n=1}^r \int_0^t \cdots \int_0^{s_{n-1}} c_n(t, s_1, \dots, s_n) E[h(s_1) \cdots h(s_n)] ds_n \cdots ds_1 \quad . \end{aligned} \quad (3.18)$$

However,

$$\begin{aligned} E_0\{c(t)E_0[L_t|F_t^Y]\} &= E_0\{c(t) \frac{dP}{dP_0}\} = Ec(t) \\ &= c_0(t) + \sum_{n=1}^r \int_0^t \cdots \int_0^{s_{n-1}} c_n(t, s_1, \dots, s_n) E[h(s_1) \cdots h(s_n)] ds_n \cdots ds_1 \quad . \end{aligned} \quad (3.19)$$

by lemma 2.2. Applying (3.18) and (3.19) in (3.17) yields

$$E_0\{c(t)E_0[L_t^{(r)}|F_t^Y]\} = 0 \quad .$$

(3.16) is established in analogous fashion using a version of lemma 2.2 for expressions

$f(t) I_t^n(c_n)$, $n < r$, under the condition $E\{f^2(t)(\int_0^T h^2(s)ds)^r\} < \infty$.

Proof of theorem 3.2 Because of lemma 3.3 it suffices to show (3.14) holds if and only if

$$E[\tilde{f}(t) - \hat{f}(t)]^2 < E[c(t) - \hat{f}(t)]^2$$

for all $c(t) \in Y_x(t)$. Since

$$E[c(t) - \hat{f}(t)]^2 = E[c(t) - \tilde{f}(t)]^2 + E[\tilde{f}(t) - f(t)]^2 + 2E[c(t) - \tilde{f}(t)](\tilde{f}(t) - \hat{f}(t))$$

this will occur if and only if

$$E[c(t) - \tilde{f}(t)](\tilde{f}(t) - f(t)) = 0 \quad \forall c(t) \in Y_x(t). \quad (3.20)$$

Thus, we will demonstrate (3.20). Begin by noting that

$$E\left[\frac{dP_0}{dP_t} | F_t^Y\right] = (E_0\left[\frac{dP}{dP_0} | F_t^Y\right])^{-1} = (E_0[L_t | F_t^Y])^{-1}.$$

Then

$$\begin{aligned} E[c(t) - \tilde{f}(t)](\tilde{f}(t) - \hat{f}(t)) &= E\left\{\frac{(c(t) - \tilde{f}(t))(\tilde{f}(t)E_0[L_t | F_t^Y] - E_0[f(t)L_t | F_t^Y])}{E_0[L_t | F_t^Y]}\right\} \\ &= E\left\{E\left[\frac{dP_0}{dP_t} | F_t^Y\right](c(t) - \tilde{f}(t))(\tilde{f}(t)E_0[L_t | F_t^Y] - E_0[f(t)L_t | F_t^Y])\right\} \quad (3.21) \\ &= E_0\{(c(t) - \tilde{f}(t))(\tilde{f}(t)E_0[L_t | F_t^Y] - E_0[f(t)L_t | F_t^Y])\}. \end{aligned}$$

Next note from (3.13) that $g_j(t)$ depends on kernels k_n of at most order $j+r$. Thus

$$\begin{aligned} \tilde{f}(t)E_0[L_t | F_t^Y] &= \left[\sum_{j=0}^r I_t^j(g_j(t)) \right] \left[\sum_{j=0}^{2r} I_t^j(k_j) + E_0^{L(2r)} | F_t^Y \right] \\ &= \sum_{j=0}^r I_t^j(g_j(t)) + \sum_{j=r+1}^{3r} I_t^j(\bar{g}_j(t)) + \tilde{f}(t)E_0^{L(2r)} | F_t^Y \end{aligned}$$

where the $\bar{g}_j(t)$, $r+1 < j < 3r$ are determined by the multiplication formula. Using the partial expansions of theorem 3.1, we then see that the expression $\tilde{f}(t)E_0[L_t | F_t^Y] - E_0[f(t)L_t | F_t^Y]$ appearing in (3.21) equals

$$\begin{aligned}
& \sum_{j=0}^r I_t^j (g_j(t) - l_j(t)) + \sum_{j=r+1}^{2r} I_t^j (\bar{g}_j(t) - l_j(t)) \\
& + \sum_{j=2r+1}^{3r} I_t^j (\bar{g}_j(t)) + \tilde{f}(t) E_0 \{ L_t^{(2r)} | F_t^y \} \\
& - E_0 \{ f(t) L_t^{(2r)} | F_t^y \} .
\end{aligned} \tag{3.22}$$

Since $y(\cdot)$ is Brownian on (Ω, P_0) , multiple integrals of different orders are orthogonal on (Ω, P_0) , and so if (3.22) is used in (3.21) we find

$$\begin{aligned}
(3.21) = & [c_0(t) - a_0(t)] [g_0(t) - l_0(t)] + \\
& + \sum_{j=1}^r \int_0^t \cdots \int_0^{s_{j-1}} (c_j - a_j)(g_j - l_j) ds_j \cdots ds_1 \\
& + E_0 \{ (c(t) - \tilde{f}(t)) \tilde{f}(t) E_0 \{ L_t^{(2r)} | F_t^y \} \} \\
& - E_0 \{ (c(t) - \tilde{f}(t)) E_0 \{ f(t) L_t^{(2r)} | F_t^y \} \} ,
\end{aligned} \tag{3.23}$$

The last two terms of (3.23) are zero by lemma 3.3. Thus, it is clear that (3.23), and hence (3.20), is zero iff

$$g_j = l_j \quad 0 \leq j \leq r .$$

This completes the proof.

The technique of theorem 4.2 extends to other problems as well. Suppose, for instance, that a filter

$$a'(t) = a'_0(t) + \sum_{j=1}^q I_t^j (a'_j(t))$$

of order q is available; $a'(t)$ need not be the best q th order filter. Let $r > q$, and, rather than ask for the best r th order filter, let us seek the "best r th order correction" to $a'(t)$, i.e., the mean-square minimizing $a(t)$ of the form

$$a(t) = a'(t) + \sum_{j=q+1}^r I_t^j(a_j(t))$$

where $a_j(t)$, $j = q+1, \dots, r$ are free to be chosen. Define the kernels $g_j(t)$ as before, but with $a_j(t)$ replaced by $a'_j(t)$ for $0 < j < q$.

Theorem 3.3. Let the hypotheses of theorem 4.2 hold. Then $a(t)$ is the best r th order correction to $a'(t)$ if and only if

$$g_j(t, s_1, \dots, s_j) = E\{f(t)h(s_1)\dots h(s_j)\}, \quad q+1 < j < r. \quad (3.24)$$

Proof. As before, it suffices to show that (3.24) holds iff

$$E[c(t)-a(t)][a(t)-\hat{f}(t)] = 0$$

for all $c(t) = a'(t) + \sum_{j=q+1}^r I_t^j(c_j(t))$. By the same calculations as in theorem 4.2

$$E[c(t)-a(t)][a(t)-\hat{f}(t)]$$

$$\begin{aligned} &= E_0[(c(t)-a(t))[a(t)E_0\{L_t|F_t^Y\} - E_0\{f(t)L_t|F_t^Y\}]] \\ &= E_0\left\{\sum_{j=q+1}^r I_t^j(c_j-a_j)\left(\sum_{j=0}^r I_t^j(g_j(t)-I_j(t)) + \sum_{j=r+1}^{2r} I_t^j(\bar{g}_j-I_j)\right.\right. \\ &\quad \left.\left.+ \sum_{j=2r+1}^{3r} I_t^j(g_j(t)+a(t)E_0[L_t^{(2r)}|F_t^Y] - E_0\{f(t)L_t^{(2r)}|F_t^Y\})\right)\right\} \\ &= \sum_{j=q+1}^r \int_0^t \dots \int_0^{s_{j-1}} (c_j-a_j)(t, s_1, \dots, s_j) (g_j-I_j)(t, s_1, \dots, s_j) ds_j \dots ds_1. \end{aligned}$$

This equals zero iff $g_j = I_j$ for $q+1 < j < r$.

Remark Clearly, an analogous result holds for the case in which an arbitrary subset of $\{a_j\}_{j=0}^r$ is given and the remainder are chosen as to optimize the mean-square filter error. Thus, if a_j , $j \in \{j_1, \dots, j_q\} \subseteq \{0, 1, \dots, r\}$ are given, then the $\{a_j(t)\}$, $j \notin \{j_1, \dots, j_q\}$ are optimally chosen iff $g_j = I_j$ for every $j \in \{0, 1, \dots, r\} - \{j_1, \dots, j_q\}$.

As a first example of theorem 3.2 let us compute the kernel equations for the best linear estimate $\tilde{f}(t) = a_0(t) + \int_0^t a_1(t,s)dy(s)$. From (3.13),

$$g_0(t) = a_0(t) + \int_0^t a_1(t, \sigma) E[h(\sigma)] d\sigma$$

$$g_1(t, s) = a_1(t, s) + \int_0^t a_1(t, \sigma) E[h(s)h(\sigma)] d\sigma + a_0(t) Eh(s) .$$

The kernel equations are then

$$a_0(t) + \int_0^t a_1(t, \sigma) Eh(\sigma) d\sigma = Ef(t)$$

$$a_0(t) Eh(s) = a_1(t, s) + \int_0^t a_1(t, \sigma) E[h(s)h(\sigma)] d\sigma = Ef(t)h(s) ,$$

or, eliminating $a_0(t)$ from the second equation,

$$\begin{aligned} a_0(t) + \int_0^t a_1(t, \sigma) E[h(\sigma)] d\sigma &= Ef(t) \\ a_1(t, s) + \int_0^t a_1(t, \sigma) \text{cov}(h(s), h(\sigma)) d\sigma &= \text{cov}[f(t), h(s)] \end{aligned} \quad (3.25)$$

(3.25) is, of course, the well-known Wiener-Hopf type equation for optimal linear filtering. Before examining higher order examples, we will discuss the Kalman filter.

3.3 The Kalman filter

Consider the filtering problem in which $h(t, x) = H(t)x$ and $x(t)$ is a Gauss-Markov process arising as the solution of the system

$$dx(t) = F(t)x(t)dt + G(t)db(t)$$

where $x_0 = \text{constant or a Gaussian m.v. independent of the Brownian motion } b(\cdot)$. The celebrated Kalman-Bucy theorem states that the optimal state estimator $\hat{x}(t) = E\{x(t) | F_t^Y\}$ satisfies the equation

$$\begin{aligned} d\hat{x}(t) &= F(t)\hat{x}(t)dt + P(t)H^T(t)(dy(t) - H(t)\hat{x}(t)dt) \\ \hat{x}(0) &= x_0 \end{aligned} \quad (3.26)$$

where $P(t)$ is the solution of a deterministic Riccati equations. It follows that $\hat{x}(t)$ is, in fact, a linear functional of $y(\cdot)$; if $\Phi(t,s)$ denotes the state transition matrix of $F(t) - P(t)H^T(t)H(t)$, the solution to (3.26) is

$$\hat{x}(t) = \Phi(t,0)x_0 + \int_0^t \Phi(t,s)P(s)H^T(s)dy(s) . \quad (3.27)$$

This simple, linear structure is not an immediate consequence of the expansion formulae of theorem 3.1, because, even in this case, both numerator and denominator series will be truly infinite sums. It is therefore of interest to see how $\hat{x}(t)$ can be derived from the general expansion. We will show that this can be done using theorem 3.2 and moment equalities for Gaussian random variables.

The most common proof of the Kalman-Bucy filter invokes the stochastic differential equation for the conditional moments (cf. Fujisaki, Kallianpur, Kunita [2]). In this approach, the equation for $x(t)$ requires knowledge of $\hat{x}^2(t)$, that for $\hat{x}^2(t)$ knowledge of $\hat{x}^3(t)$, and so on, thus leading to an infinite, coupled set of equations. To derive the Kalman-Bucy theorem, it must be independently argued that the conditional distribution of $x(t)$ given \mathcal{F}_t^y is Gaussian. Because of identities between different moments of Gaussian m.v.'s, this allows the moment equations to be truncated at $n = 2$ and leads to (3.26) and (3.27). By way of contrast, the derivation here will not require explicitly knowing the conditional density. For other methods of deriving the Kalman-Bucy filter, see Van Schuppen [20].

In the interest of computational simplicity, we will consider only the most simple case:

$$\begin{aligned} dx(t) &= db(t) \quad x(0) = 0 \\ dy(t) &= x(t)dt + dw(t) \quad y(0) = 0 , \end{aligned} \quad (3.28)$$

where $b(\cdot)$ and $w(\cdot)$ are independent, standard Brownian motions. The techniques work also for the general case.

Theorem 3.4 $\hat{x}(t) = \int_0^t a(t,s)dy(s)$ where $a(t,s)$ satisfies the Wiener-Hopf equation

$$a(t,s) + \int_0^t a(t,\sigma) \min(s,\sigma)d\sigma = s \quad t > s .$$

Before presenting the proof we must recall the following moment identities (Miller [16], Marcus-Willsky [14]).

Lemma 3.5. Let $[z_1, \dots, z_k]$ be a jointly Gaussian random vector. Then

$$E[z_1, \dots, z_k] = Ez_1Ez_2 \cdots z_k + \sum_{j=2}^k \text{cov}[z_1, z_j] E[\prod_{l \neq j} z_l] .$$

Proof of theorem 3.4. Since $y(\cdot)$ is continuous and Gaussian, the set of polynomials in $y(\cdot)$ is dense in $L^2(\Omega, F_t^Y, P)$, (Kallianpur [8]). Therefore, it suffices to show that $\int_0^t a(t,s)dy(s)$ is the best r th order estimate for every r , $1 \leq r < \infty$. Since

$$E(\int_0^t b^2(s)ds)^r < \infty$$

$$E b^2(t)(\int_0^t b^2(s)ds)^r < \infty$$

for all r and t , theorem 4.2 applies. That is, if $g_j(t, \dots)$, $0 \leq j < \infty$ are defined so that

$$\int_0^t a(t,s)dy(s) \sum_{i=0}^{\infty} I_t^i(k_i) = \sum_{i=0}^{\infty} I_t^i(g_i)$$

$\int_0^t a(t,s)dy(s)$ is the best r th order estimate if and only if

$$g_j(t, s_1, \dots, s_j) = E(b(t)b(s_1)\cdots b(s_j)) \quad 0 \leq j < r .$$

From (3.14), we may easily calculate

$$\left. \begin{aligned} g_0(t) &= 0 \\ g_j(t, \dots) &= j(a(t, \cdot) \Theta(t) k_{j-1})(\dots) \\ &\quad + (a(t, \cdot) \Theta_1(t) k_{j+1})(\dots) \quad j > 0 . \end{aligned} \right\} \quad (3.29)$$

However

$$\begin{aligned}
 & j(a(t, \cdot) \otimes(t) k_{j-1})(s_1, \dots, s_j) \\
 &= \frac{1}{j!} \sum_{\pi \in S_j} a(t, s_{\pi(1)}) E \left\{ \prod_{l=2}^j b(s_{\pi(l)}) \right\} \\
 &= \sum_{i=1}^j a(t, s_i) E \left\{ \prod_{\substack{l \neq i \\ l \leq j}} b(s_l) \right\}
 \end{aligned}$$

and

$$(a(t, \cdot) \otimes_1(t) k_{j+1})(s_1, \dots, s_j) = \int_0^t a(t, \sigma) E \{ b(\sigma) b(s_1) \dots b(s_j) \} d\sigma . \quad (3.30)$$

The kernel equations (3.29) become

$$0 = E b(t) \quad (3.31)$$

$$a(t, s) + \int_0^t a(t, \sigma) E \{ b(\sigma) b(s) \} d\sigma = E \{ b(t) b(s) \} \quad (3.32)$$

$$\begin{aligned}
 & \sum_{i=1}^j a(t, s_i) E \left\{ \prod_{\substack{l \leq j \\ l \neq i}} b(s_l) \right\} + \int_0^t a(t, \sigma) E \left\{ b(\sigma) \prod_{\substack{l \leq j \\ l \neq i}} b(s_l) \right\} d\sigma \\
 &= E \{ b(t) b(s_1) \dots b(s_j) \} , \quad j \geq 2 \quad (3.33)_j
 \end{aligned}$$

(3.31) is true by definition, and (3.32), by hypothesis. It remains to prove that (3.33)_j, $j \geq 2$ all hold. However, a direct application of lemma 3.5 shows that

$$E \{ b(\sigma) \prod_{l=1}^j b(s_l) \} = \sum_{i=1}^j \min(\sigma, s_i) E \left\{ \prod_{\substack{l \leq j \\ l \neq i}} b(s_l) \right\}$$

for every j . Using this, the left hand side of (3.33)_j becomes

$$\begin{aligned}
& \sum_{i=1}^j \{a(t, s_i) + \int_0^t a(t, s) \min(s, s_i) ds\} E\left\{\prod_{\substack{l < j \\ l \neq i}} b(s_l)\right\} \\
& = \sum_{i=1}^j \min(s_i, t) E\left\{\prod_{\substack{l < j \\ l \neq i}} b(s_l)\right\} \\
& = E\{b(t)b(s_1)\cdots b(s_j)\}
\end{aligned}$$

where the first equality employs the hypothesis on $a(t, s)$, and the second employs lemma 3.5 again. Thus $(3.33)_j$ is true for all $j > 2$.

3.4 Quadratic Filters

As a further example of the technique of section 3.2, we will present the optimal kernel equations for the quadratic case ($r = 2$) and sketch a theoretical approach to their solution. To guarantee validity of the discussion, assume throughout the hypotheses of theorem 3.2 for $r = 2$.

Deriving the optimal kernel equations is simply a matter of calculation. Let $\tilde{f}(t) = a_0(t) + \int_0^t a_1(t, s) dy(s) + \int_0^t \int_0^{s_1} a_2(t, s_1, s_2) dy(s_2) dy(s_1)$ and let $g_j(t, \dots)$ be defined from a_0, a_1, a_2 in the manner indicated at (3.13). Thus

$$g_0(t) = a_0(t) + a_1(t) \Theta_1 k_1 + a_2(t, \cdot) \Theta_2 k_2 \quad (3.34)$$

$$\begin{aligned}
g_1(t, s) = & a_1(t, s) + a_0(t) k_1(s) + (a_1(t, \cdot) \Theta_1 k_1)(s) \\
& + (a_2(t, \cdot) \Theta_1 k_1)(s) + (a_2(t, \cdot) \Theta_2 k_3)(s)
\end{aligned} \quad (3.35)$$

$$\begin{aligned}
g_2(t, s_1, s_2) = & a_2(t, s_1, s_2) + a_0(t) k_2(s_1, s_2) + (a_1(t, \cdot) \Theta_1 k_1)(s_1, s_2) \\
& + (a_1(t, \cdot) \Theta_1 k_3)(s_1, s_2) + 2(a_2(t, \cdot) \Theta_1 k_2)(s_1, s_2) \\
& + (a_2(t, \cdot) \Theta_2 k_4)(s_1, s_2) \quad (3.36)
\end{aligned}$$

(More properly, Θ in (3.34) - (3.36) should be $\Theta(t)$.) According to theorem 3.2 $\tilde{f}(t)$ is optimal quadratic iff g_0, g_1 , and g_2 , are respectively, $Ef(t)$, $Ef(t)h(s)$ and $Ef(t)h(s_1)h(s_2)$. Recalling the definition of $\Theta(t)$ from section 2 and $k_j = Eh(s_1)\dots h(s_j)$, we derive for the optimal kernel equations:

$$Ef(t) = a_0(t) + \int_0^t a_1(t,s)Eh(s)ds + \int_0^t \int_0^{s_1} a_2(t,s_1,s_2)Eh(s_1)h(s_2)ds_2 ds_1 \quad (3.37)$$

$$Ef(t)h(s) = a_1(t,s) + a_0(t)Eh(s) + \int_0^t a_1(t,\sigma)Eh(\sigma)h(s)d\sigma$$

$$+ \int_0^t a_2(t,s,\sigma)Eh(\sigma)d\sigma + \int_0^t \int_0^{\sigma_1} a_2(t,\sigma_1,\sigma_2)Eh(\sigma_1)h(\sigma_2)h(s)ds \quad (3.38)$$

$$Ef(t)h(s_1)h(s_2) = a_2(t,s_1,s_2) + a_0(t)Eh(s_1)h(s_2) + a_1(t,s_1)Eh(s_2)$$

$$+ a_1(t,s_2)Eh(s_1) + \int_0^t a_1(t,\sigma)Eh(\sigma)h(s_1)h(s_2)d\sigma \quad (3.39)$$

$$+ \int_0^t [a_2(t,s_1,\sigma)Eh(\sigma)h(s_2) + a_2(t,s_2,\sigma)Eh(\sigma)h(s_1)]d\sigma$$

$$+ \int_0^t \int_0^{\sigma_1} a_2(t,\sigma_1,\sigma_2)E\{h(s_1)h(s_2)h(\sigma_1)h(\sigma_2)\}d\sigma_2 d\sigma_1 .$$

These equations deserve some elementary remarks before we set about solving them.

First, the optimal kernels are all interrelated in the general case. We cannot solve for a_0 and a_1 independently of knowing a_2 . Likewise, if $a_0 = c_0$, $a_1 = c_1$ are the kernels of the best linear estimate, they will not, in general, be the lower order kernels of the best quadratic estimate. Secondly, the equation (3.37) - (3.39) can be used for other suboptimal designs in the spirit of theorem 3.3. Thus, if a_0 and a_1 are given, and we seek the best quadratic correction to $a_0(t) + \int_0^t a_1(t,s)dy(s)$, this will be found by solving (3.39) for a_2 in terms of a_1 and a_0 . The methods developed for solving the full set of equations will also apply to the best correction problem.

As a system of integral equations, (3.37) - (3.39) looks complicated and contains unusual features. Nevertheless, we will show that solving the system can be reduced to two, familiar tasks -- solving a linear estimation problem and solving a Fredholm integral equation. The method behind this reduction is simply to eliminate a_0 and a_1 to obtain an equation for a_2 . The basic steps are: 1) eliminate $a_0(t)$ from (3.38) to derive the integral equation (3.41) for a_1 ; 2) solve this for a_1 in terms of a_2 using the solution to the linear filtering problem, (see 3.42); 3) use (3.42) to eliminate a_1 from

(3.39) and derive (3.43), an integral equation only involving the unknown a_2 , and; 4) turn (3.43) into the Fredholm equation (3.45). The central equation is thus (3.45). Once it is solved for a_2 , a_1 and a_0 are found by using (3.42) and (3.37) respectively.

Let $R : L^2([0,t]) \rightarrow L^2([0,t])$ be the operator defined by

$$(R\beta)(s) = \int_0^t \text{cov}[h(s), h(\sigma)] \beta(\sigma) d\sigma . \quad (3.40)$$

The first step is easy; simply solve (3.37) for $a_0(t)$ and substitute the result in (3.38). We then derive

$$\begin{aligned} [I + R]a_1(t, \cdot)(s) &= \text{cov}[f(t), h(s)] - \int_0^t Eh(\sigma)a_2(t, s, \sigma)d\sigma \\ &\quad - \int_0^t \int_0^{s_1} \text{cov}[h(s), h(\sigma_1)h(\sigma_2)]a_2(t, \sigma_1, \sigma_2)d\sigma_2 d\sigma_1 . \end{aligned} \quad (3.41)$$

The next step, solving this for a_1 , thus requires inverting $I + R$.

Lemma 3.6

i) $h(s)$, $s < t$, has a best linear estimate $\bar{h}(s) = a_0(s) + \int_0^s a(s, \sigma)v(\sigma)d\sigma$

(As a convention, set $a(s, \sigma) = 0$ for $0 < s < \sigma < t$)

ii) $[I + R]^{-1} = I - Q$ where Q is the integral operator with kernel

$$q(s_1, s_2) = a(s_1, s_2) + a(s_2, s_1) - \int_0^t a(\sigma, s_1)a(\sigma, s_2)d\sigma$$

$$0 < s_1, s_2 < t .$$

Proof. We are assuming $E[\int_0^t h^2(s)ds]^4 < \infty$. This guarantees that $\bar{h}(s)$ exists, and, as in (3.25),

$$a(s_1, s_2) + \int_0^{s_1} a(s_1, \sigma)\text{cov}[h(s_2), h(\sigma)]d\sigma = \text{cov}[h(s_1), h(s_2)]$$

$$0 < s_2 < s_1 < t .$$

ii) is standard. See, for instance, Gessey [3].

This lemma can now be applied to solve (3.41) for $a_1(t, s)$:

$$\begin{aligned} a_1(t, s) &= \text{cov}[f(t), h(s)] - \int_0^t q(s, \sigma)\text{cov}[f(t), h(\sigma)]d\sigma \\ &\quad - \int_0^t \int_0^t r'(t, s, \sigma_1, \sigma_2)a_2(t, \sigma_1, \sigma_2)d\sigma_2 d\sigma_1 \end{aligned} \quad (3.42)$$

where

$$\begin{aligned} r'(t, s, \sigma_1, \sigma_2) &= \frac{1}{2} \operatorname{cov}[h(s), h(\sigma_1)h(\sigma_2)] \\ &+ \frac{1}{2} [q(s, \sigma_2)Eh(\sigma_1) + q(s, \sigma_1)Eh(\sigma_2)] \\ &+ \frac{1}{2} \int_0^t q(s, \sigma) \operatorname{cov}[h(\sigma), h(\sigma_1)h(\sigma_2)] d\sigma . \end{aligned}$$

In deriving r' , advantage was taken of the (assumed) symmetry of $a_2(t, s_1, s_2)$ in s_1 , s_2 . Now, using (3.37) and (3.42), we may eliminate a_0 and a_1 from equation (3.39).

The result is

$$\begin{aligned} a_2(t, s_1, s_2) &= F(t, s_1, s_2) \\ &- \int_0^t [r_1(s_1, \sigma)a_2(t, s_2, \sigma) + r_1(s_2, \sigma)a_2(t, s_1, \sigma)] d\sigma \\ &- \int_0^t \int_0^t r_2(t, s_1, s_2, \sigma_1, \sigma_2)a_2(t, \sigma_1, \sigma_2) d\sigma_2 d\sigma_1 \end{aligned} \quad (3.43)$$

where

$$\begin{aligned} F(t, s_1, s_2) &= \operatorname{cov}[f(t), h(s_1)h(s_2)] \\ &- \int_0^t \operatorname{cov}[h(s_1), h(s_2)h(\sigma)] (\operatorname{cov}[f(t), h(\sigma)]) \\ &- \int_0^t q(\sigma_1, \sigma_2) \operatorname{cov}[f(t), h(\sigma_2)] d\sigma_2 d\sigma \end{aligned}$$

$$r_1(s, \sigma) = \operatorname{cov}[h(s), h(\sigma)]$$

$$\begin{aligned} r_2(t, s_1, s_2, \sigma_1, \sigma_2) &= \frac{1}{2} [\operatorname{cov}[h(s_1), h(s_2), h(\sigma_1)h(\sigma_2)] \\ &- \operatorname{cov}[h(s_1), h(s_2)] \operatorname{cov}[h(\sigma_1), h(\sigma_2)]] \\ &- \int_0^t \operatorname{cov}[h(s_1), h(s_2), h(n)] r'(t, n, \sigma_1, \sigma_2) dn . \end{aligned}$$

It remains to solve (3.43) for a_2 . This is simply a linear integral equation for a_2 . However, its middle term, involving a tensor contraction between a_2 and r_1 , is not standard, and the usual linear integration theory does not apply directly. Despite this,

it is possible to rewrite (3.43) as a Fredholm integral equation and thereby to reduce the task of calculating a_2 to a familiar problem. First notice that (3.43) may be rewritten in the form

$$a_2(s_1, s_2) = F(s_1, s_2) - (Ra_2(s_2, \cdot))(s_1) - (Ra_2(s_1, \cdot))(s_2)$$

$$- \int_0^t \int_0^t r_2(s_1, s_2, \sigma_1, \sigma_2) a_2(\sigma_1, \sigma_2) d\sigma_2 d\sigma_1$$

or

$$[(I + R)a_2(s_2, \cdot)](s_1) = F(s_1, s_2) - (Ra_2(s_1, \cdot))(s_2)$$

$$- \int_0^t \int_0^t r_2(s_1, s_2, \sigma_1, \sigma_2) a_2(\sigma_1, \sigma_2) d\sigma_2 d\sigma_1 . \quad (3.44)$$

In these equations, the argument t has been omitted for simplicity. Now apply $(I + R)^{-1}$ to both sides of (3.44). Again, an equation of the form

$$[(I + R)a_2(s_1, \cdot)](s_2) = \text{linear terms in } a_2$$

is obtained, but this time there are no partial tensor contractions of the form $Ra(s_1, \cdot)(s_2)$ on the right hand side. With a final application $(I + R)^{-1} = I - Q$ to both sides the following Fredholm equation for a_2 is derived.

$$a_2(t, s_1, s_2) = F_1(t, s_1, s_2) + \int_0^t \int_0^t \gamma(t, \sigma_1, s_2, \sigma_1, \sigma_2) a_2(t, \sigma_1, \sigma_2) d\sigma_2 d\sigma_1 \quad (3.45)$$

where

$$F_1(t, s_1, s_2) = F(t, s_1, s_2) - \int_0^t [q(s_2, \sigma_1) F(t, s_1, \sigma_2) + q(s_1, \sigma_2) F(t, \sigma_1, s_2)] d\sigma_2 d\sigma_1$$

$$+ \int_0^t \int_0^t q(s_1, \sigma_1) q(s_2, \sigma_2) F(t, \sigma_1, \sigma_2) d\sigma_2 d\sigma_1$$

$$\gamma(t, s_1, s_2, \sigma_1, \sigma_2) = \gamma_1(t, s_1, s_2, \sigma_1, \sigma_2) - \int_0^t q(s, u) \gamma_1(u, s_2, \sigma_1, \sigma_2) du$$

$$\gamma_1(t, s_1, s_2, \sigma_1, \sigma_2) = -r_2(t, s_1, s_2, \sigma_1, \sigma_2) - q(s_2, \sigma_1) q(s_1, \sigma_2)$$

$$+ \int_0^t q(s_2, u) r_2(t, s_1, u, \sigma_1, \sigma_2) du .$$

Remarks. The viewpoint here is not recursive. Rather t is fixed throughout and integral operators are defined and inverted on $L^2([0,t])$ or $L^2([0,t]^2)$, and at a later time t the whole operation would have to be repeated. This poses an interesting question for further research. What structure on the moments $Eh(s)$, $Ef(t)h(s)$, etc., would allow a recursive solution to the quadratic kernel equations, in the sense that $a(t+dt, s_1, s_2)$ could be constructed in a simple way from $a(t, s_1, s_2)$? A related question is also important. When are the solutions a_1 and a_2 separable functions? If separability occurred, then, as mentioned above, the stochastic integrals $I_t^1(a_1)$ and $I_t^2(a_2)$ could be realized as the outputs of stochastic differential systems. Certainly, if F and γ of the Fredholm equation for a_2 are separable, a_2 will be separable, but due to the complicated manner in which the moments $Ef(t)$, $Ef(t)h(s)$, etc., combine to produce F and γ , this does not lead to easy conditions. This issue is not pursued further.

REFERENCES

1. J. S. Eterno, Nonlinear Estimation and Phase-Lock Loops, Ph.D. Dissertation, Dept. of Aeronautics and Astronautics, 1976, MIT.
2. M. Fujisaki, G. Kallianpur, H. Kunita, Stochastic Differential Equations for the Nonlinear Filtering Problem, Osaka J. Math. 9 (1972), 19-40.
3. R. Gessey, Canonical Representations of Second Order Processes with Applications, Ph.D. Dissertation, Dept. of Electrical Engineering, Stanford University, 1969.
4. T. Hida, Generalized Multiple Wiener Integrals, Proc. Japan Acad. Ser. A Math. Sci., 54 (1978), no. 3, 55-58.
5. T. Hida, Stationary Stochastic Processes, Princeton University Press, Princeton, NJ, 1970.
6. T. Hida and G. Kallianpu, The Square of a Gaussian Markov Process and Nonlinear Prediction, J. of Multivariate Anal., 5 (1977), 451-461.
7. K. Ito, Multiple Wiener Integrals, J. Math Soc. Japan 3 (1951), 157-169.
8. G. Kallianpur, The Role of Reproducing Kernel Hilbert Spaces in the Study of Gaussian Processes, in Advances in Probability, P. Ney, ed., Marcel Dekker, New York, 1970, 49-83.
9. G. Kallianpur and C. Striebel, Estimation of Stochastic Systems: Arbitrary System Process with Additive White Noise Observation Errors, Ann. Math. Stat. 39 (1968), 785-801.
10. R. E. Kalman and R. S. Bucy, "New Results in Linear Filtering and Prediction Theory, J. Basic Engr. (Trans. ASME), Vol. 83-D, 1961, 95-108.
11. R. S. Liptser and A. N. Shirayev, Statistics of Random Processes I, Springer-Verlag, New York, 1977.
12. W. Magnus and F. Oberhettinger, Formulas and Theorems for the Special Functions of Mathematical Physics, Springer-Verlag, New York, 1966.
13. S. I. Marcus, S. K. Mitter and D. Ocone, Finite Dimensional Nonlinear Estimation in Continuous and Discrete Time, Analysis and optimization of stochastic systems, (Proc. Internat. Conf., Univ. Oxford, Oxford, 1978), pp 419-436, Academic Press, London, 1980.
14. S. I. Marcus and A. S. Willsky, Algebraic Structure and Finite Dimensional Nonlinear Estimation, SIAM J. Math. Anal. 9(2), 1978, 312-325.
15. P. A. Meyer, Un cours sur les intégrales stochastiques, in Séminaire de Probabilités X, Springer Lecture Notes 511, Springer-Verlag, 1976.
16. K. S. Miller, Multidimensional Gaussian Distributions, Wiley, 1965.
17. S. K. Mitter and D. Ocone, Multiple Integral Expansions for Nonlinear Filtering, Proceedings of the 18th IEEE Conference on Decision and Control, Ft. Lauderdale, Fl, 1979.

18. D. Ocone, Topics in Nonlinear Filtering Theory, Ph.D. Thesis, Department of Mathematics, MIT, 1980.
19. M. Reed and B. Simon, Methods of Mathematical Physics II: Fourier Analysis, Self Adjointness, Academic Press, New York, 1975.
20. J. H. Van Schuppen, Notes on the Stochastic Filtering Problem and the Stochastic Realization Problem, Part I, Instituto Matematico <<G. Castelnuovo>>, Universita degli Studi Roma, 1979.
21. E. Wong, Stochastic Processes in Information and Dynamical Systems, McGraw-Hill, 1972.
22. E. Wong, Explicit Solutions to a Class of Nonlinear Filtering Problems, to appear in Stochastics.

DO/jvs

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER #2226	2. GOVT ACCESSION NO. <i>AD-A103 873</i>	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Multiple Integral Expansions for Nonlinear Filtering	5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period	
7. AUTHOR(s) Daniel Ocone	6. PERFORMING ORG. REPORT NUMBER	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin 610 Walnut Street Madison, Wisconsin 53706	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Work Unit Number 4 - Statistics & Probability	
11. CONTROLLING OFFICE NAME AND ADDRESS See Item 18 below	12. REPORT DATE June 1981	
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	13. NUMBER OF PAGES 43	
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		15. SECURITY CLASS. (of this report) UNCLASSIFIED
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES U.S. Army Research Office Air Force Office of Scientific Research National Science Foundation P. O. Box 12211 Washington, D.C. 20550 Research Triangle Park Washington, D.C. 20332 North Carolina 27709		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Multiple stochastic integral, nonlinear filtering		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Multiple stochastic integral expansions are applied to the problem of filtering a signal observed in additive noise. It is shown that the optimal mean-square estimate may be represented as a ratio of two multiple integral series. A formula for expanding the product of two multiple integrals is developed and applied to deriving equations for the kernels of best, finite expansion approximations to the optimal filter. These equations are studied in detail in the quadratic case.		